

PCTWorld Intellectual Property Organization
International BureauINTERNATIONAL APPLICATION PUBLISHED IN ACCORDANCE
WITH THE PATENT COOPERATION TREATY (PCT)

| | | |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| (51) International Patent Classification⁶: C07H 21/00, C07K 1/04, C12N 15/10, C12P 21/00 | A1 | (11) International Publication Number: WO 95/17413 (43) International Publication Date: June 29, 1995 (6/29/95) |
| (21) International Application Number: PCT/EP94/04240 (22) International Filing Date: December 20, 1994 (12/20/94) (30) Priority Data: P 43 43 591.2 December 21, 1993 (12/21/93) DE (71) Applicant (for all Designated States except US): EVOTEC BIOSYSTEMS GMBH [DE/DE]; Grandweg 64, D-22529 Hamburg (DE). (72) Inventor: and (75) Inventor/Applicant (for US only): SCHWIENHORST, Andreas [DE/DE]; Weender Landstrasse 68, D-37075 Göttingen (DE). LINDEMANN, Björn [DE/DE]; Tannenhof 115b, D-22397 Hamburg (DE). FUCHS, Merle [DE/DE]; Stiegel 5, D-37077 Göttingen (DE). HENCO, Karsten [DE/DE]; Kirchberg 4, D-40699 Erkrath (DE). (74) Agent: MEYERS, Hans-Wilhelm et al.; Deichmannhaus am Hauptbahnhof, D-50667 Cologne (DE). | | (81) Designated States: JP, US, European Patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published: <i>With International Search Report.</i> |
| (54) Title: METHOD FOR THE EVOLUTIVE DESIGN AND SYNTHESIS OF FUNCTIONAL POLYMERS BASED ON SHAPE ELEMENTS AND SHAPE CODES (57) Abstract: A method for preparing oligomeric or polymeric functional elements from shape elements, in which the functional elements are obtainable by linking at least two shape elements, at least one of which is itself composed of at least two monomers that are linked by at least one chemical bond which corresponds to the chemical bond between two shape elements. | | |

Method for the evolutive design and synthesis of functional polymers based on shape elements and shape codes

The subject matter of the present invention is a method according to Claim 1.

The rapid development in the field of biosciences in recent years has stimulated not only basic research, but also applied research in this field. In this regard proteins play a major role due to their broad activity spectrum. An entire branch of modern biotechnology is presently concerned with protein engineering, i.e., the production of designer proteins which are developed either on the basis of known proteins by gradual modification, or by completely new synthesis. A distinction is made primarily between two approaches, rational and irrational design.

The aim of rational design is to produce an amino acid sequence which folds into a desired structure and which also has the desired function. Therefore, this strategy obviously depends on a thorough understanding of protein folding. Recent advances have involved, among other things, the rational design of simple structure domains. However, the design of larger proteins having complex or even unparalleled new properties is still beyond the capabilities of this approach. On the other hand, irrational design does not require information about the protein structure, protein folding, etc. The only prerequisite in this regard is knowledge of the desired property and the ability to evaluate molecular populations with respect to this property. Based on a combinatorial library composed of peptides or

proteins, molecules having the desired property are selected, and analyzed only in retrospect. Thus, in this case the mechanism by which a molecule achieves the given task is not determined in advance.

Although this approach has quite recently produced in a very elegant manner peptides having simple and sometimes new properties, here as well there is the problem of how to obtain larger proteins having more complex functions. A complete 20mer bank having $20^{20} = 10^{26}$ different sequences provides an astronomical number of molecules for investigation. If the peptide sequence is also to be encoded by a nucleic acid, the problem is even more serious. Because the genetic code is degenerated, i.e., an amino acid may be represented by several different codons, this results in at least $4^{60} = 10^{36}$ molecules that are synthesized. Normally, in order to largely avoid stop codons, only G or C is allowed at the third codon position. The remaining number of 10^{30} molecules still exceeds the standard yield from commercial DNA synthesis by 12 orders of magnitude. A further reduction in the codons allowed at each position has been proposed by Youvan. It remains to be seen whether this method will not unacceptably limit the measurable sequence space, in particular in the search for new functions.

Nature uses modular systems to construct functional structures. Nucleotide building blocks, amino acid building blocks (encoded as nucleotide triplets), and exon domains (built from amino acid building blocks) are known. The evolutive optimization of functional biopolymers according to patent application WO 92/18645 is based on the concept of finding an optimal structure by continuous improvement of existing base properties of an enzyme, for example, in the continuous adaptation to desired reaction conditions such as ionic strength, temperature, and pH. If advantageous or at least neutral mutations are possible, by multiple repetitions of selection and mutation even remote regions of the

sequence space are accessible which did not belong to the starting population. However, in no step of this method is a structure obtained that is different from the original, already functional structure. A property of the starting molecule is optimized which is already inherent in the original molecule, even if only to a minor extent. The "path" that such an evolution takes through the sequence space is determined by the available ridges in the underlying value system leading in the direction of the optima. The same as for all methods which do not fully make the sequence space available, in this procedure there is the practically inestimable risk of becoming "stuck" in a local optimum. In practice, this means that certain regions of the sequence space, including the optima located there, are separated by wide, deep "valleys." However, for the limited population size of molecular species in experiments (P 43 22 147, WO 92/18645), the probability is too low of producing remote multiple-error mutants which are located on the other side of this barrier and which indicate the route to these new optima.

Nature has developed a number of mechanisms to deal with this problem: long development time frames, recombination methods (horizontal gene transfer, crossing over, gene conversion, exon recombination (exon shuffling), virus shuttles, mobile elements (transposons), subunit structure of complex proteins) as well as multigene families containing pseudogenes.

Due to the number of mutant formation and selection processes conducted in parallel, this may increase the chance of producing a desired multiple-error mutant; mutated gene segments may be efficiently mixed by recombination. Functionless pseudogenes, as members of a functional multigene family, may be maintained in existence as multiple-error mutants for fairly long development times, without counterselection, in order to once again be positively selectable with possible retention of a function.

Transfer of these mechanisms to an efficient in vitro optimization is obviously not straightforward. However, in each case the difficulties must be addressed for such tasks for which continuous optimization cannot be expected. This applies in particular for adaptation processes in which a function must be established completely anew.

The technical object on which the invention is based relates to providing a method for preparing oligomeric or polymeric functional elements, such as biopolymers, having functional properties, for example enzymes, ribozymes, active substances, etc. The aim is to provide a method which is superior to the conventional screening methods and which makes use of evolutive strategies.

This object is achieved by a method having the features of Claim 1. The subsequent subclaims concern preferred embodiments of the method according to the invention.

According to the invention, in order to prepare oligomeric or polymeric functional elements from shape elements, shape elements are first built by chemical or enzymatic linkage of at least two monomers, and the shape elements obtainable in this manner are then linked to form functional elements. The type of chemical bonding between the monomers corresponds to that between the respective shape elements. The functional elements obtainable in this manner may then be tested for the specific potential functions. The advantages of the procedure according to the invention are illustrated by the following description.

The shape elements are preferably linked using a solid phase as reaction support. The linkage of the shape elements may be carried out by chemical and/or

enzymatic means. The linkage of the shape elements to the functional elements may be performed either systematically by targeted addition of the individual shape elements and subsequent linkage, or statistically by randomly controlled addition of the functional elements and linkage thereof. It is possible to carry out the linkage in a step-by-step stereospecific and/or targeted manner.

Nucleic acids, double-stranded or single-stranded DNA and/or RNA, and/or modified nucleic acids are preferred as shape elements. Peptides and/or polypeptides and/or other chemical oligomer shape elements capable of coupling are suitable as shape elements. Oligosaccharides or polysaccharides may also be included.

In one preferred embodiment of the method according to the invention, the shape elements are used as already synthesized oligomer building blocks, or are prepared in the reaction vessel in situ, in a manner of speaking.

It is advantageous to carry out the reaction of the shape elements in parallel microreaction assays (as proposed in P 43 22 147.5) in which the shape elements are linked in a predetermined sequence. In particular, after synthesis is completed the reaction products, such as functional elements or precursors thereof, remain bound to the solid phase, and after separation of the reactants are further processed or removed from the solid phase. However, the reaction may also be carried out in solution under suitable reaction conditions known to one skilled in the art, or the reactions, which are carried out coupled to the solid phase or in homogenous solution, may be combined with one another.

Use of fluorescence correlation spectroscopy (FCS) (PCT/EP 93/01291) allows the mode of operation of the functional elements to be directly evaluated in the

same volume element in which the synthesis proceeds. This involves a very direct option for monitoring the result of a developing functional element synthesis.

For the stepwise linkage of the shape element, in each case a shape element as reactant is coupled to the solid phase for each reaction step. Mixtures of shape elements may also be used, and/or may be generated directly in the reaction vessel. When nucleic acids are used as shape elements, it is advantageous to provide at least one reactant with an interface of a restriction enzyme, or to use a nucleic acid shape element that is free of start codons and/or stop codons. The reaction interfaces are preferably those which may be recognized by class IIS restriction enzymes. Introduction of restriction interfaces of this enzyme class is advantageous, since any given sequences may be selectively linked without the choice of the reaction enzyme influencing the sequence requirements for the end product.

If single-stranded overhangs are introduced into the shape elements to be linked, any given sequences may be selectively linked via same without having to impose any requirements on the sequence of the desired end product. This requirement may also be met by selective and reversible chemical and/or enzymatic modification of the 3' and/or 5' ends of the nucleic acids, for example by phosphorylation instead of and in combination with the introduction of the single-stranded overhangs.

The coupling of a trityl protective group which may be cleaved by treatment with acetic acid is named as one example of a reversible chemical modification. Introduction of the trityl group at the 3' or 5' end of the nucleotide results in blockage of the ligation of the shape codes and/or function codes or functional elements.

A 3' or 5' end may be modified by treating an oligonucleotide or polynucleotide with nuclease; for example, the 3' end is modified by digestion by treatment with exonuclease III. If nucleotide triphosphates are introduced into the corresponding oligonucleotide or polynucleotide (DNA, for example), the exonuclease stops the digestion at the first thionucleotide. This results in a regulatable modification of the end of the oligonucleotide or polynucleotide.

The method according to the invention allows the use of shape elements which are known from natural function domains of proteins and polypeptides according to X-ray crystallographic analysis. In this manner, already known building blocks or modules of functional elements occurring in nature may be used.

The shape elements to be used may also be obtained from selection experiments.

It is particularly advantageous to use shape elements in a length of 1 to 60 amino acids or nucleotide sequences having a corresponding encoding length. The shape elements may also be degenerated at certain positions and/or may bear deletions or insertions, in particular when nucleotides are used as shape elements.

Also claimed is use of the method according to the invention as described above for synthesis of design libraries, developed in parallel, of functional oligomers or polymers.

The original task of combinatorial libraries is primarily to offer a variety of functions as a variety of sequences. It is presently taken for granted that the three-dimensional structure of proteins is relatively stable with respect to substitutions of individual amino acids. Due to the large number of elucidated protein

structures it is known that, although proteins may have little or no sequence homology, they are still able to assume the same or very similar three-dimensional structure. This may possibly be based on the fact that only a limited number of possible folding modes of amino acid chains are stable under biological conditions. However, structural relatedness also reflects the evolution of recent proteins from a relatively limited number of original structures or modules. These modules may be understood as small functional domains or compact structural units, and may also be easily detected in present genes. In the "exon shuffling" hypothesis it is presumed that the evolution to more complex proteins has been tremendously accelerated specifically by the combination of exons, i.e., modules in the sense described above. If it is assumed that the number of exons to be identified which would allow building of all currently known proteins is between 1000 and 7000, a hierarchical strategy of protein design, using building blocks of increasing complexity, opens up the possibility for much more rapid measurement of a shape space with an associated fitness landscape than would be permitted in a search using a traditional combinatorial library. According to conventional methods, a protein composed of 150 amino acids (the size of a classical nucleotide binding site, the so-called "Rossmann fold") would have to be selected from a library of $20^{150} = 10^{195}$ different amino acid sequences. In contrast, combinations of 1000 different modules having a length of 30 amino acids result in a complexity of only $1000^5 = 10^{15}$ molecules.

The method according to the invention is a hierarchical method for designing proteins, nucleic acids and derivatives thereof, or chemical oligopolymers or polymers having certain desired properties, based on module libraries, referred to below as "shape elements." According to the invention, the shape elements may also be gene segments which encode for shape elements. The shape elements which function as modules should be randomly combinable. Smaller proteins or

subunits for larger proteins having certain properties are sorted from the pool of module combinations in a subsequent selection step, and may be reused as building blocks in a subunit library, and so forth.

As the result of faulty copying of individual building blocks, "noise" may also be introduced at the amino acid sequence level at each building stage. This allows modulation of the three-dimensional configuration of chemical groups, and therefore further functional optimization of selected molecules. The proposed strategy requires a new type of artificial gene assembly.

Heretofore, primarily two methods have been used, having the common feature that the DNA is ligated in a certain orientation in order to identify the sequence of the amino acids. Probably the oldest method, developed by Khorana et al., uses overlapping complementary single-stranded DNA molecules which are hybridized with one another before the ligation. The second method utilizes interfaces of restriction enzymes in the gene to be built, in order to subdivide the gene at these locations into blocks which are then combined in multiple successive steps. By use of both methods, the sequence at the transitions of the oligo-DNAs or blocks used are identified in a methodical manner. However, this does not exactly meet the requirement for arbitrary exchangeability of the individual modules in the building phase of the gene. Thus, a component of the present invention by necessity is also a new type of artificial gene assembly. According to the invention, the following general procedure is followed:

- The method for artificial gene assembly operates similarly to the method described in WO 92/18645;
- The method enables operation not with the variance in the sequence space,

– 10 –

but, rather, with the variance in the so-called shape space. The shape space, formed from base elements of defined stable shape elements, reduces the complexity of the variants of the components of the sequence space;

- The method makes the function space available via variation of building blocks of the shape space;
- Building blocks of the shape code (see below) are used as building blocks;
- For selection of the building blocks, certain selection criteria are used for preselection which correspond to theoretical assumptions or natural shape analogs.

As modules for variation (mutation) and selection conducted in parallel, heretofore only the nucleotides or amino acids have been available as synthetically or enzymatically manageable building blocks of a polymer for targeted coupling processes. As described above, in many cases direct access to a functional surface structure of a polymer is denied due to the problem of the large numbers of variants of the sequence space.

The object of this evolutive adaptation process is the use of modular building blocks, the shape code, composed of the shape elements. The shape code includes shape elements built from elements of the sequence space. The shape code, which may be derived, for example, from natural polymers such as proteins, polypeptides, or functional nucleic acids, encodes stable shape elements under fixed external conditions (secondary structures, possibly containing tertiary structural elements). It is noteworthy that very different sequences (primary structures) may be encoded for very similar shape elements. In other words,

elements that are very close to one another in the shape space may be widely separated from one another in the sequence space (large Hamming distance). The same applies for the reverse case. Within the meaning of the invention, this very property illustrates that the exchange of sequences in the sequence space which are the same with regard to design may signify a great step in terms of a multiple-error mutant. This requirement may be technically implemented with the aid of the synthesis method addressed by the invention. The corresponding subdivisions are produced by programmed synthesis, but are not achieved by defective replication in the sense of faulty PCR methods, as described in WO 92/18645.

For the linear combination of shape codes, heterologous as well as at least partially homologous shape codes may be used which are genotypically defined by a shape code of natural or artificial origin.

According to the invention, "natural origin" is understood to mean that genetic information that is already present, such as that stored in the genome of organisms, for example, is relied on. According to the invention, "shape codes of artificial origin," in particular at the nucleic acid level, is understood to mean that sequences may be generated by algorithms, using data processing systems, in order to be subsequently chemically synthesized according to these instructions. Lastly, preparation by de novo synthesis is also possible by reacting polymerases with the associated substrates such as nucleotides. The polymerase reaction may be carried out in a matrix-dependent or matrix-independent manner.

The shape elements and functional elements as understood in particular according to the invention may be construed as phenotypes, for example in proteins or peptides. The corresponding genotypes, for example at the nucleic acid level, are the shape codes and function codes which correspond to same. If one remains at

the nucleic acid level, for example, the "phenotype" having functional elements and/or shape elements is embodied by a ribozyme, for example, which is correspondingly reflected genotypically in a nucleic acid sequence as shape code and/or function code. This means that according to the invention the terms "functional element/shape element" (phenotype) are always understood, in a manner of speaking, as complementary to the terms "function code/shape code" (genotype).

Provided that they are nucleic acids, the shape elements and/or functional elements or shape codes and/or function codes may be obtained by various methods as described above, namely, by using nucleic acid sequences that are already known, by generating synthetic sequences in data processing systems, or by de novo synthesis.

Figure 8 illustrates the terms sequence space, shape space, and function space. Similarly as for the considered relationship of shape space and sequence space, it also applies for the relationship of shape space and function space that closely adjacent, homologous elements in the shape space may be widely separated from one another in the function space. As schematically indicated in Figure 8, the geometry and the physicochemical topology and dynamics of the molecular surface which interacts with a second molecule is crucial for the function of a polymer. The underlying structure, defined based on the shape code, could have a very different chemical nature. Similar functions in the function space are illustrated by similar boundary surface topologies.

In particular with regard to the relatively small molecular populations that may be achieved in experiments, it is of critical importance that the produced variation in the shape space represents the possible functional variety in the function space in

a much more efficient manner than, for example, the variation in the sequence space.

The invention is schematically explained in greater detail in the following description of the figures, with reference to examples.

Figure 1 concerns two single-stranded DNA or RNA molecules which are chemically or enzymatically ligated (T4 RNA ligase, for example), one of the molecules being immobilized on the solid phase via a cleavable linker (biotin streptavidin, for example), and the other molecule being freely present in solution.

Numerous solid phase materials (magnetic surface-activated plastic beads, for example) are currently available for this purpose. This method allows the stepwise development of larger DNAs or RNAs. After each ligation step unreacted RNAs are flushed away, and the ligation products present on the solid phase are transferred in the next ligation assay. The handling, in particular the purification of the particular ligation products, is advantageously very simple.

After conclusion of the last ligation, the product is directly used as effector molecule, or in an (in vitro) translation reaction is initially translated into the corresponding protein structure, which then functions as an effector molecule.

Figure 2 concerns two completely double-stranded DNA molecules which are subjected to chemical or enzymatic "blunt end" ligation (T4 DNA ligase, for example), one of the DNA molecules being immobilized on the solid phase via a cleavable linker, and the other DNA molecule being freely present in solution. In this manner, larger double-stranded DNA molecules may be developed step-by-step. The targeted ligation is achieved by differing phosphorylation of the

reactants. Module A and the last module are designed in such a way that they each contain an interface for a restriction enzyme. This allows, firstly, cleavage of the product from the solid phase, and secondly, subsequent targeted cloning of the DNA (also see Figure 5).

Figure 3: According to Figure 2, DNA molecules may also be ligated when the molecule which is present in solution has a single-stranded end on one side, i.e., is not present in completely double-stranded form. This end is therefore not available for the double strand-specific ligation using T4 DNA ligase, for example. In combination with the previously mentioned phosphorylation strategies (Figure 2, in particular variant 1), it is possible to carry out the ligation without undesired by-products. The DNA molecule present in solution may be designed in such a way that, preceding its single-stranded end, it still has the interface of a restriction enzyme, preferably that of a Class IIS enzyme (AlwI, for example), having a recognition site in the partially single-stranded DNA segment to be cleaved. After the ligation, the ligation product on the solid phase may be cleaved using the restriction enzyme. This once again results in a completely double-stranded DNA molecule on the solid phase. Alternatively, the single-stranded end may be replenished to form a double strand, using a polymerase, or may be digested using an exonuclease.

Figure 4: Restriction interfaces (overlapping) may also be formed by ligating two double-stranded DNA molecules with one another.

Figure 5: Completely or partially double-stranded DNA molecules may be ligated according to Figures 1 through 4, even when mixtures of molecules (B, C, D, for example) are used. This results in mixtures of immobilized molecules, each corresponding to different combinations of the building blocks used. At the end of

the last ligation step the total DNA or a portion thereof may be cleaved from the solid phase, using restriction enzymes which cleave within the construct, and optionally cloned in a phage or bacterial display system. However, the DNA may also be expressed in a combined in vitro transcription and translation system.

Figure 6: Based on module libraries, it is possible to produce peptides, protein domains, and small proteins by random combination of individual modules. Corresponding to a hierarchical method for protein design, in another step protein domains may also be combined as building blocks. Mutations may be inserted at any level of complexity which, without altering the global structure, allow fine-tuning of the three-dimensional configuration of chemical groups.

Figure 7 schematically illustrates that various proteins have homologous functions (chymotrypsin/trypsin) despite different catalytically active amino acids in the active center in relation to the substrate, or are able to catalyze totally different reactions (trypsin/elastase) despite a similar spatial configuration of the amino acids in the active center.

Figure 8 illustrates the relationship of the terms sequence space, shape space, and function space.

Figures 9 through 11:

Oligomers or polymers are prepared by matrix-dependent enzymatic or chemical synthesis by extension of stochastic (randomized) or selected (constructed) primer molecules. The primers may be discrete sequences which are complementary to 1) exclusively at the end of the original matrix molecule (Figure 9/10), or overall discrete sequences in the original matrix molecule (Figure 10/11), composed of a mixture of random sequences which randomly allow initiation of the synthesis at

many locations, depending on the (partial) complementarity (Figure 10).

Either the primers or, as shown in the figure, the matrix DNA may be biotinylated for simplifying subsequent purification procedures. This would be particularly advantageous for the strand separation (on streptavidin Dynabeads, for example) for purifying the extended primers.

Figure 12

Instead of standard dNTPs (deoxynucleoside triphosphates), thio-NTPs in particular are used. Chain-terminating molecules may then be standard dideoxynucleoside triphosphates (ddNTPs), for example. It is known that phosphodiester bonds may be easily cleaved by exonuclease III in 50 mM tris/HCl and 5 mM MgCl₂, at pH 10.0, in a period of minutes. In contrast, thiophosphate bonds are not cleaved (Labeit et al., DNA 5:173, 1986). In this manner, after the polymerase is inactivated the ends of the resulting polymers may be "deprotected" by enzymatic removal of the ddNTPs.

Figure 13

The resulting deprotected polymers may be removed by either thermal means or chemical means (using NaOH, for example), whereby the biotin-coupled molecules may be separated on streptavidin Dynabeads, for example. After physiological conditions are set, the deprotected polymers hybridize to form partially overlapping duplexes.

Figure 14

PCR without primer results in actual (re)combination of the polymers and further extension thereof. After the partially overlapping duplexes are completed, further PCR with (end-position) primers takes place, resulting in products which once

again have the original length. Sequences in which multiple markers are combined or recombined are also included.

The sequence space is defined by the linear neighbor relationships of the polymer components of a polymer structure. Homologies describe similarities (in %) in the sequence of the components of a chemical substance class. The greater the degree of relatedness of two sequences, the smaller the distance in the sequence space.

a) ... AATAATGCGCAATATTAGGCCT ...
b) ... AATAAAAAGCAATATTAAGCCT ...
c) ... TTAGCTAGCGATGCGCGCCGGG ...

For example, the sequences a) and b) have significant homology, whereas sequence c) shows no similarities at all with a) and b).

The shape space is defined by the "spatial" neighbor relationships of the polymers which they represent. The distance between two sequences is determined by the degree of relatedness of their structures. In the present context, "homology" means similarity of the overall structures of polymers, which once again are composed of chemically linked components. Neighboring molecules in the shape space may be located quite far apart from one another in the sequence space, and vice versa. (Similarly as above, structure a) three alpha helices, structured; b) two alpha helices plus unstructured region with end-position short helix; c) antiparallel beta sheet composed of four leafs.) The function space is defined by the geometric, dynamic, and physical/chemical surface structure which is able to specifically interact with another molecule. Homologies describe similarities in the surface structure and the associated interaction properties.

Linear combinations of shape codes are described below by means of in vitro recombination of shape codes based on natural muteins or muteins produced in vitro.

Heterologous as well as at least partially homologous shape codes may be used for the linear combination of shape codes. Sequence-homologous shape codes may be randomly present in a mixture to be recombined, or may be intentionally selected. This mixture may contain, for example, mutant collectives of a starting sequence homologously produced from one another, as described in Eigen & Henco, WO 92/18645, or homologous genes of related or different organisms. In this regard sequences that are similar with regard to their function codes, for example, may be very different.

For various host systems, nature has evolved similar or molecularly different enzymes for the same or very similar reactions, from which it may be assumed that they provide optimal solutions for the particular environment for which they have been adapted. One example of such is penicillin acylase. This enzyme is of key importance for commercial application in the field of synthesis. Prior to a synthesis of semisynthetic derivatives of the penicillin base body, naturally synthesized penicillin must first undergo a protective acylase reaction before it may be reacylated with synthetic derivatives in a reversal of the process. Penicillin acylase may be used for both reactions. Certain reaction conditions and substrates are desirable for this reaction. However, these conditions differ from the in vivo situation of the microorganism from which the particular penicillin acylases have been isolated. This applies, for example, for the optimal position of the equilibrium for the acylated synthesis product or for the hydrolytic cleavage. It is desirable to optimize the enzyme based on the reaction number under the best-suited commercial conditions.

In particular, starting with a gene of a given naturally occurring acylase, this gene may be subjected to consecutive mutation and selection cycles. If various active mutants are found, the different variants which are selected as positive may be intermixed another time via recombination, based on the particular point mutations that are selected. The variants which are mixed in this manner may be obtained as described in WO 92/18645. Nature frequently already has a collection of positively selected muteins in the form of the enzyme genes from various microorganisms which, for example, catalyze the desired type of reaction. Based on these collections, spectra of recombined shape codes and function codes may be produced before possible further mutation cycles or a combination of mutation/recombination cycles are run through again. At the beginning of such a process, it is definitely advantageous to start with the most extensive shape codes possible, whose mutations have proven to be positive or neutral in a given context of the particular gene.

The in vitro recombination is preferably carried out according to two different strategies.

Recombination may be a generally undesired by-product of an amplification reaction in the sense of a PCR reaction. When in the saturation phase of a PCR reaction the solution is depleted of reagents or enzyme after running through many cycles and the reaction proceeds below the K_m value for certain nucleotide triphosphates, this necessarily results in incomplete synthesis products. Such events have been discussed as undesired artifacts by Simon Wain-Hobson for describing HIV variants as possible artifacts following successful PCR amplification. However, according to the invention this effect is used and controlled, and in particular is intensified, in such a way that incompletely

synthesized products become dominant as desired. When at the same time the primer-induced new synthesis is incomplete, primarily incomplete synthesis products hybridize with completely or also incompletely synthesized counterstrands. This results in molecular recombination events in which various gene segments are recombined with one another in the sense of a recombination of shape codes.

According to the invention, using another specific procedure the recombination may be controlled during, and not just after, a PCR reaction. In this regard, short oligomers are added to the standard PCR reaction which act as PCR primers only when the synthesis is initiated by thermostable polymerase at comparatively low temperatures. A standard temperature cycle is carried out whenever the correct PCR reaction is intended to dominate. If internal initiating reactions are intended, several cycles at low temperature are initiated, possibly with addition of polymerases such as DNA polymerase I, as used for oligomer-initiated labeling reactions (Sambrook, Fritsch, Maniatis, "Molecular cloning"). The resulting incomplete sequences may agglomerate in further amplification cycles, and the overhanging ends may each be replenished at the 3' end. In these reactions, according to reassociation technology of nucleic acids, known per se, it must be ensured that the sequences intended for a recombination are available in sufficient concentration to form incompletely paired duplexes within a few seconds to several minutes. To prevent strand suppression from occurring instead of a recombination event in the matrix-mediated new synthesis, in particular polymerases are used which do not induce strand suppression or which have no 5'-3' exonuclease activity. Instead, thermostable ligases may preferably be used so that recombination events are fixed by covalent linkage of the fragments.

In the method according to the invention for recombination of shape codes,

elements are used which have at least partial sequence homologies as described above. With the aid of matrix-dependent chemical or enzymatic DNA or RNA synthesis by extension of produced (randomized) primers or selected (constructed) primers, a plurality of fragments of at least one original sequence is produced (see Figures 9 through 11). Selected primers having a defined sequence may be positioned in such a way that certain regions of the DNA or RNA molecules to be processed, for example active centers, endonuclease-specific cleavage sites, or gene-regulating elements, are excluded from the recombination process and therefore remain unaltered. Use of partially randomized primers in regions of (partial) complementarity may be utilized, similarly to mutagenizing primers, for additionally initiating an increased mutation rate. By using a small, subinhibitory quantity of chain-terminating monomers in the DNA synthesis of preferably dideoxynucleotides, random termination of the extension reaction and therefore a variance in length of the synthesized polymers is achieved. The average chain length of the synthesis product may be controlled, similarly as for a sequencing reaction, via the ratio of the concentration of the termination reagent to the concentrations of the nucleotide monomers. After separation of the polymerizing agent, for example for inactivation of the enzyme, the end-position protective group, i.e., the chain-terminating monomer, may be completely or partially cleaved so that the resulting polymers are once again satisfactory substrates for the extension reaction (Figure 12). The DNA or RNA polymers deprotected in this manner are then subjected to at least one cycle of denaturation/hybridization of partially complementary strands, followed by a replenishment reaction. At the conclusion of the method, the resulting mixture of extended polymers undergoes a polymerase chain reaction, wherein the primers should preferably be situated in a complementary position with respect to the ends of the sequence originally used. This results once again in products of the original length. However, these products now contain combinations of sequence segments

of various advantageous previously selected, individual point mutations for combining in a very efficient manner instead of having to first produce them sequentially in a stochastic process.

Claims

1. Method for preparing oligomeric or polymeric functional elements from shape elements, in which the functional elements are obtainable by linking at least two shape elements, at least one of which is itself composed of at least two monomers that are linked by at least one chemical bond which corresponds to the chemical bond between two shape elements.
2. Method according to Claim 1, wherein the linkage of the shape elements is carried out linked using a solid phase as reaction support.
3. Method according to Claim 1 and/or 2, wherein the linkage of the shape elements is carried out chemically and/or enzymatically.
4. Method according to at least one of Claims 1 through 3, wherein the linkage of shape elements to functional elements is carried out systematically and/or stochastically.
5. Method according to at least one of Claims 1 through 4, wherein the linkage is carried out in a step-by-step stereospecific and/or targeted manner.
6. Method according to at least one of Claims 1 through 5, characterized in that the shape elements belong to the substance class of the nucleic acids, double-stranded and/or single-stranded DNA and/or RNA, and/or modified nucleic acids and/or peptides and/or polypeptides, and/or are developed from other chemical oligomer shape elements that are capable of coupling.
7. Method according to at least one of Claims 1 through 6, characterized in that

the shape elements are used as already synthesized oligomer building blocks, or are first generated in the reaction vessel.

8. Method according to at least one of Claims 1 through 7, characterized in that the reactions are carried out in parallel microreaction assays in which shape elements are linked in a predetermined sequence.
9. Method according to at least one of Claims 1 through 8, characterized in that after synthesis is completed the reaction products, such as functional elements or precursors thereof, remain bound to the solid phase or are removed into the soluble phase.
10. Method according to Claim 9, wherein the reaction products are combined using a biological test system, the function being measured for evaluation in the same volume element as the synthesis, for example by using the FCS analytical technique.
11. Method according to at least one of Claims 1 through 10, characterized in that in the stepwise linkage of the shape elements, in each case a shape element as reactant is coupled to the solid phase for each reaction step.
12. Method according to at least one of Claims 1 through 11, characterized in that mixtures of shape elements may be used and/or generated.
13. Method according to at least one of Claims 1 through 12, characterized in that in the case of development of nucleic acid shape elements and/or linkage of nucleic acid shape elements, at least one reactant contains an interface for a restriction enzyme and/or is free of start codons and/or stop

– 25 –

codons.

14. Method according to at least one of Claims 1 through 13, characterized in that via the introduction of restriction interfaces, in particular those for class IIS enzymes, any given sequences may be selectively linked without sequence requirements for the desired end product influencing the choice of the reaction enzyme.
15. Method according to at least one of Claims 1 through 14, characterized in that via the introduction of single-stranded overhangs and/or selective and reversible chemical and/or enzymatic modification of the 3' ends and/or the 5' ends of the nucleic acids, for example phosphorylation, any given sequences may be selectively linked without any requirements being imposed on the sequence of the desired end product.
16. Method according to at least one of Claims 1 through 15, characterized in that shape elements are used based on natural proteins or polypeptides analyzed by X-ray crystallography.
17. Method according to at least one of Claims 1 through 16, characterized in that at least one of the shape elements used originates from selection experiments.
18. Method according to at least one of Claims 1 through 17, characterized in that the shape elements contain between 1 and 60 amino acids or nucleotides having a corresponding encoding length.
19. Method according to at least one of Claims 1 through 18, characterized in

– 26 –

that shape elements are used which are degenerated at certain positions and/or which bear deletions or insertions.

20. Method according to one of Claims 1 through 19, characterized in that the functional elements and/or shape elements or the function codes and/or shape codes are deposited as oligonucleotides or polynucleotides which are obtainable by
 - generation from algorithms, in particular evolutive algorithms,
 - acceptance or modification of naturally occurring nucleic acids, and/or
 - generation by de novo synthesis of oligonucleotides/ polynucleotides by matrix-dependent or matrix-independent reactions of polymerases with nucleotides.
21. Use of the method according to at least one of Claims 1 through 20 for synthesis of design libraries, developed in parallel, of functional oligomers or polymers.

1 / 9

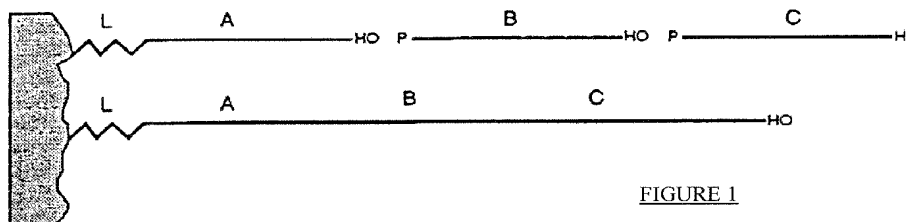


FIGURE 1

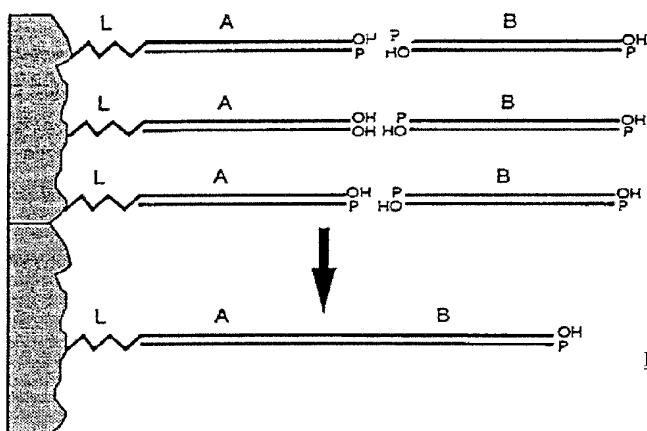


FIGURE 2

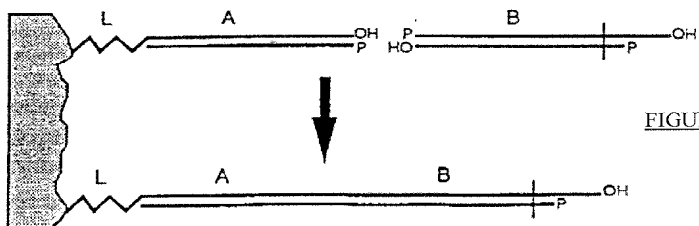


FIGURE 3

REPLACEMENT SHEET (RULE 26)

2 / 9

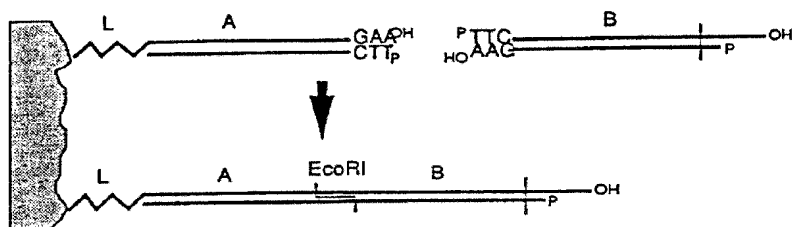


FIGURE 4

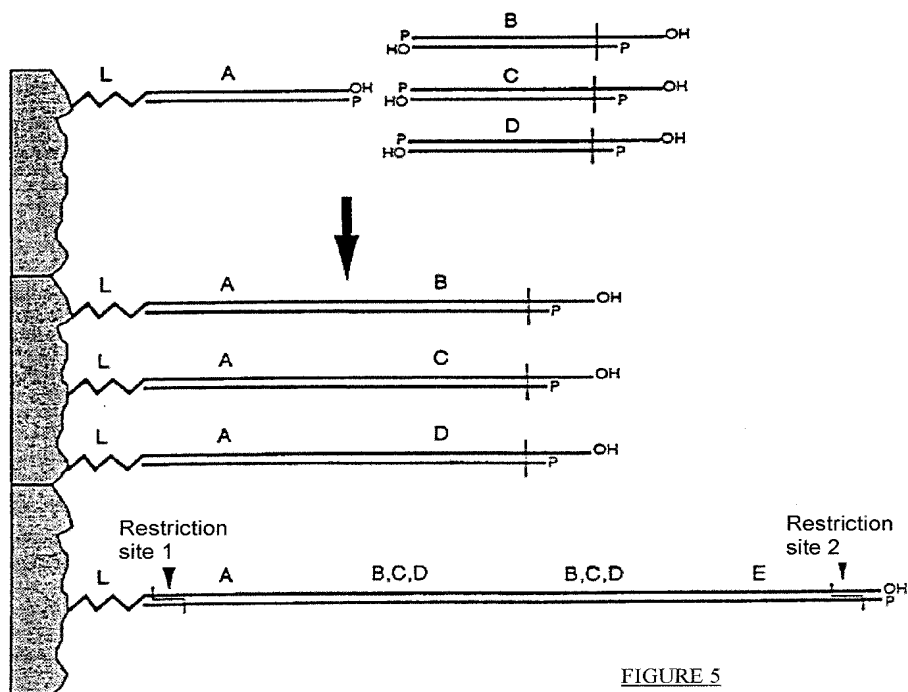
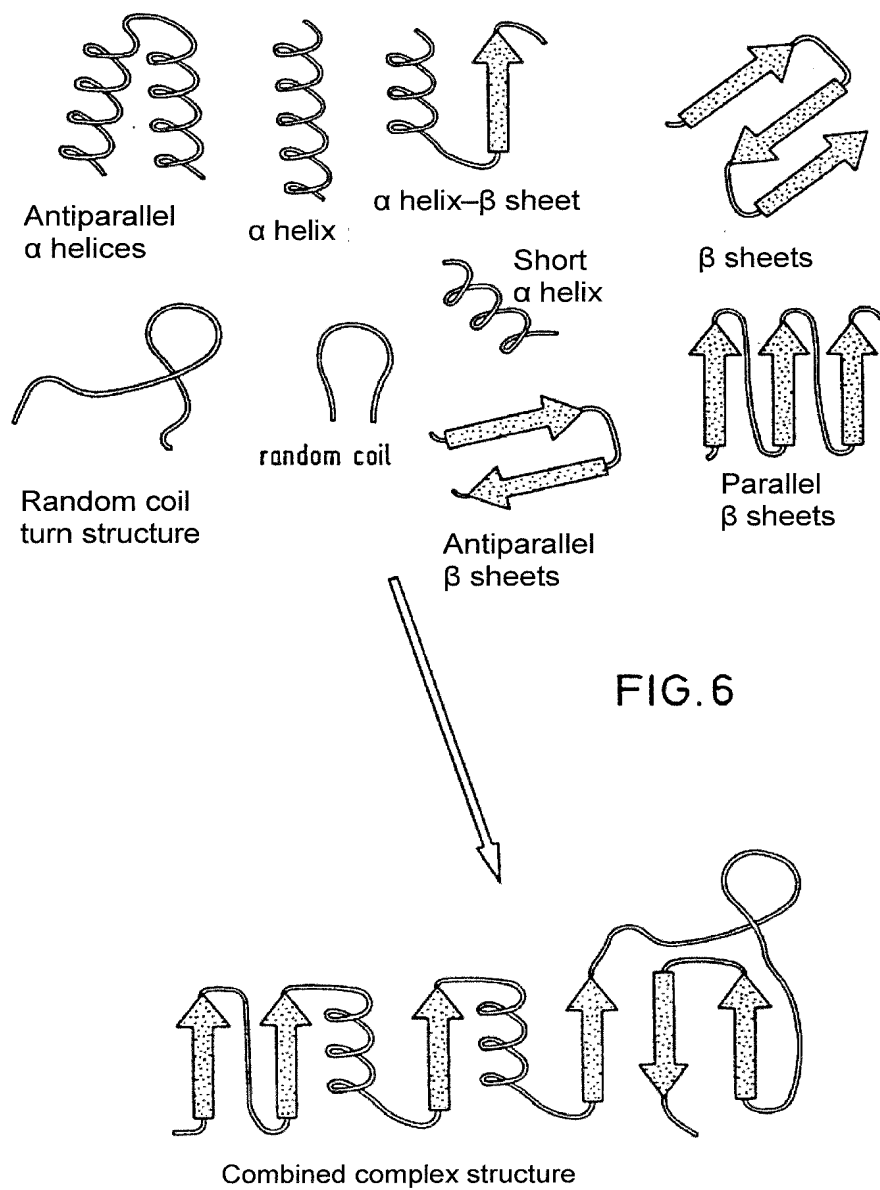


FIGURE 5



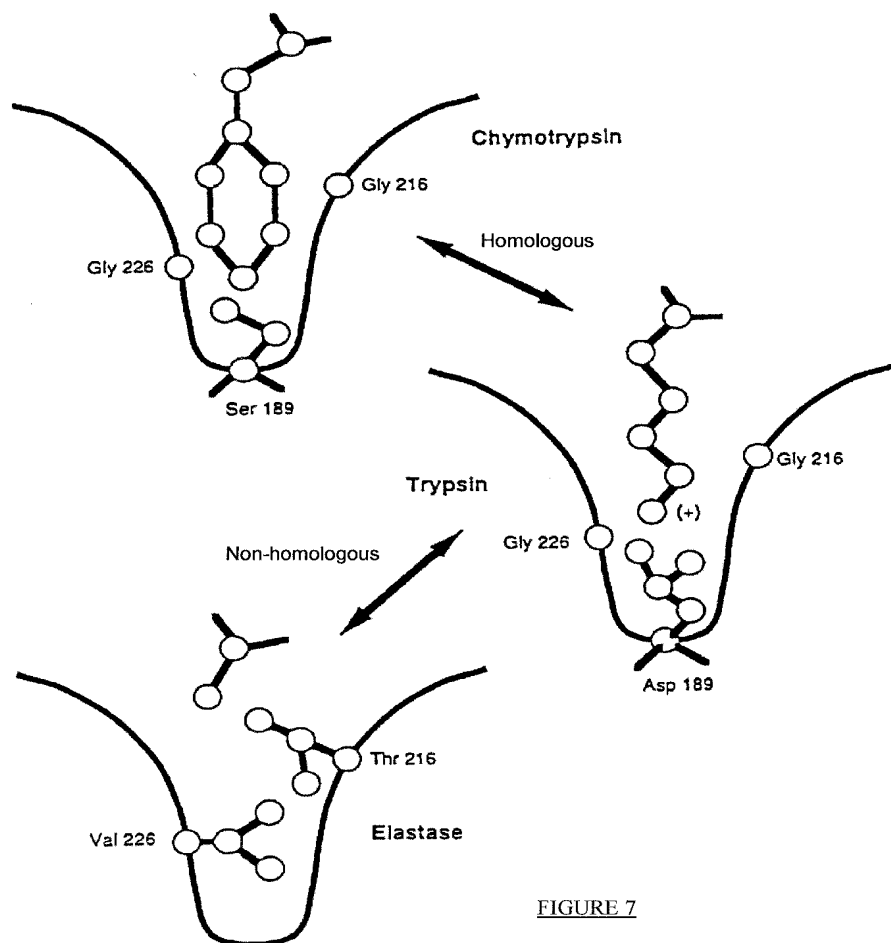
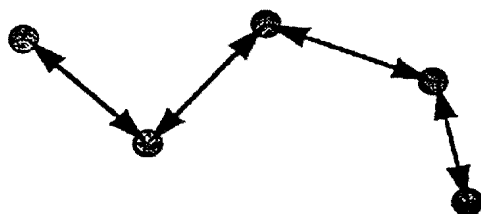
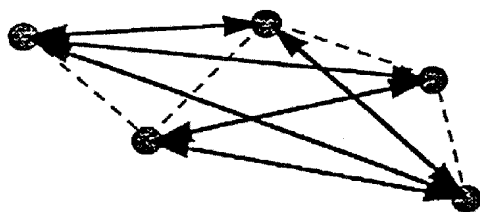


FIGURE 7

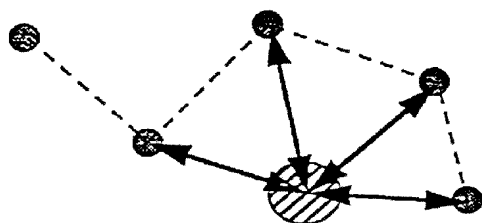
Sequence space – Shape space – Function space



The sequence space is defined by the linear neighbor relationships of a polymer structure. Homologies describe similarities (in %) in the sequence of the components of a chemical substance class.



The design space is defined by the spatial neighbor relationships of the building blocks of a polymer structure. Homologies describe similarities in the spatial structure of components which are not directly chemically linked.



The function space is defined by the geometric, dynamic, and physical/chemical surface structure which is able to specifically interact with another molecule. Homologies describe similarities in the surface structure and the associated interaction properties.

FIGURE 8

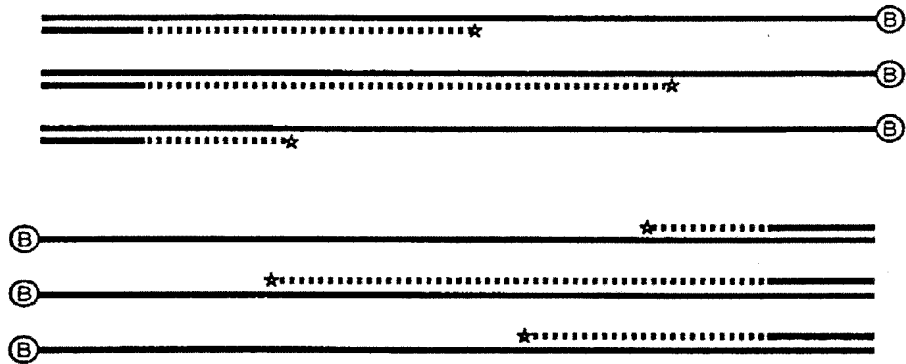


FIGURE 9

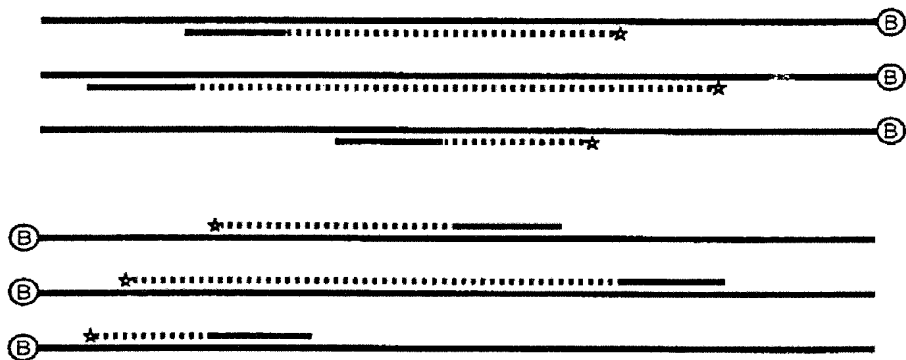


FIGURE 10

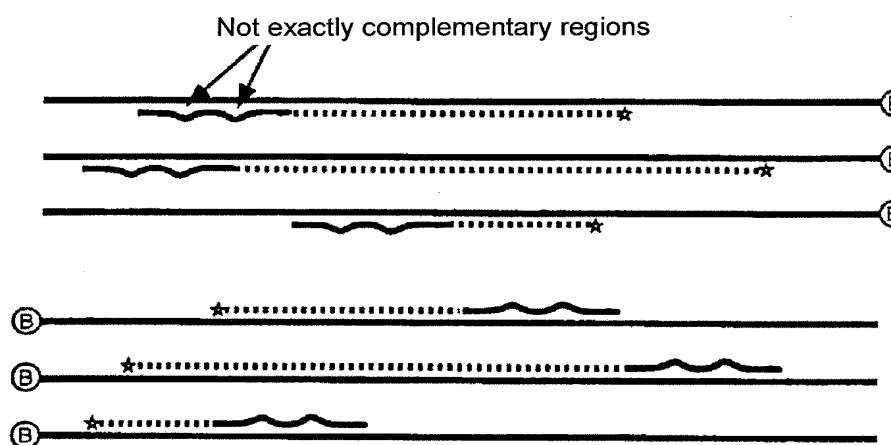


FIGURE 11

8 / 9

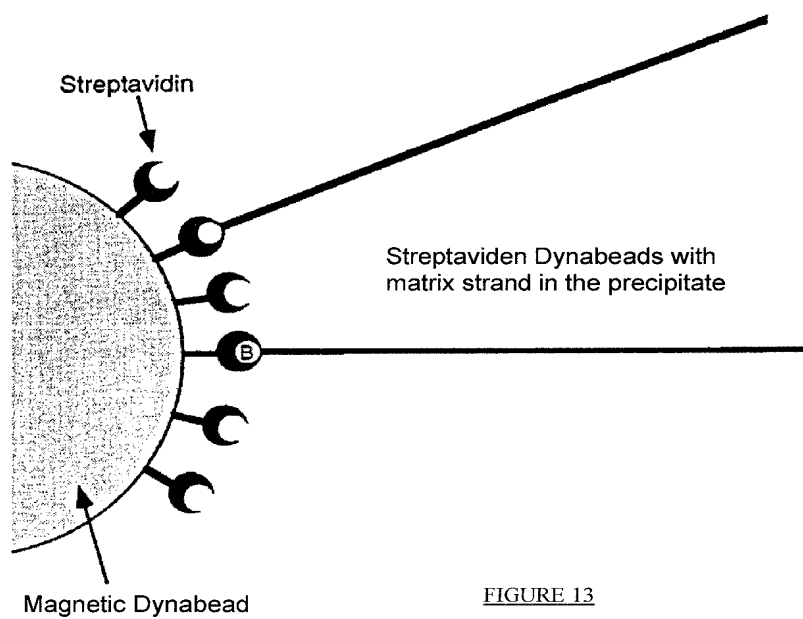
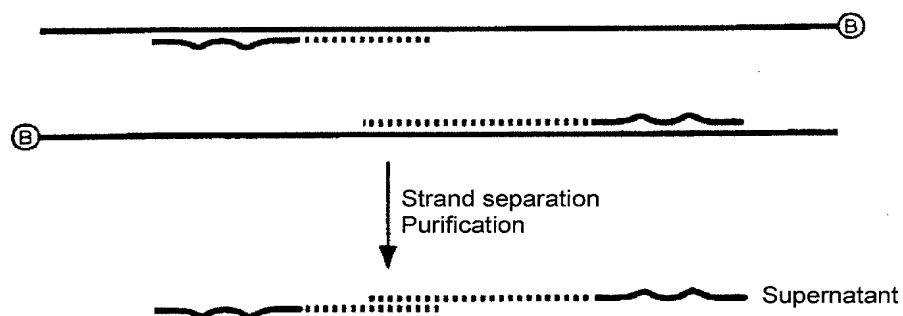


FIGURE 13

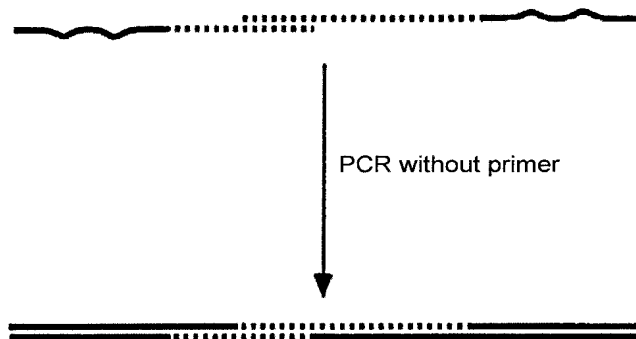


FIGURE 14